


Análise de séries temporais

por similaridade e alinhamento não
linear com Dynamic Time Warping



Mais tarefas de mineração de
séries temporais por
similaridade e o MatrixProfile



Anomalias

Anomalias

Anomalias podem ser definidas formalmente de várias maneiras.

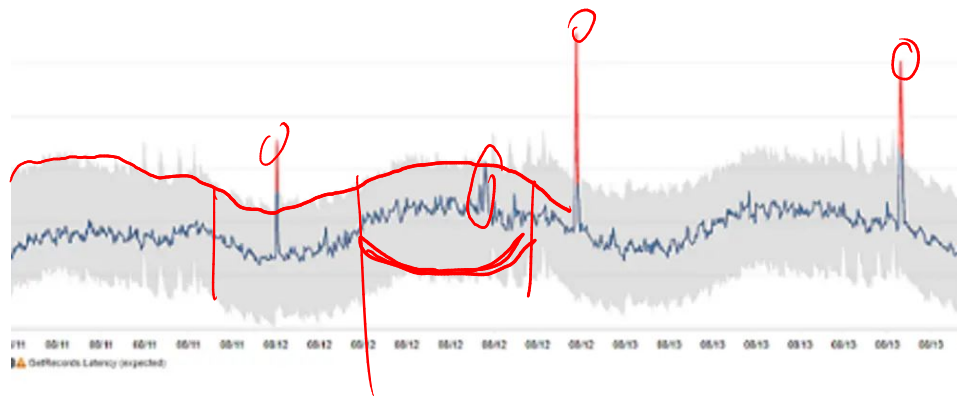
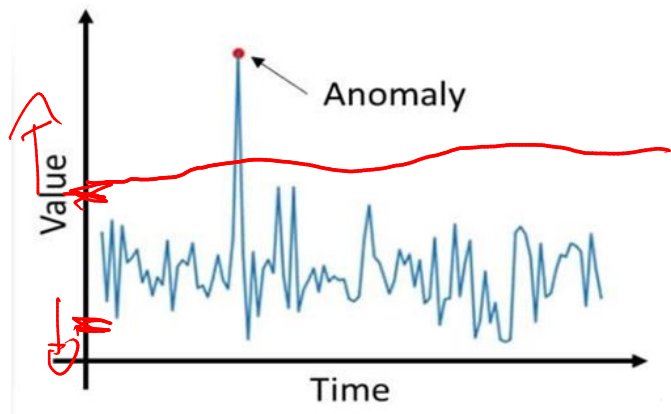
Podemos ficar com uma informal:

Anomalia: uma observação ou conjunto de observações próximas no tempo que diferem significativamente

dos demais dados

Anomalias

É muito comum vermos exemplos como esses



Mas eles são muito “óbvios”. Além disso, dependendo da aplicação, basta filtrar com mediana móvel.

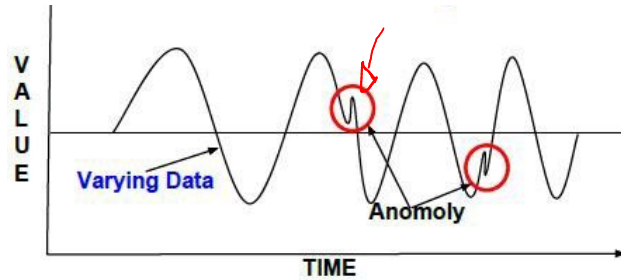
Anomalias

O cenário anterior é até comum e não pode ser ignorado. Ex:

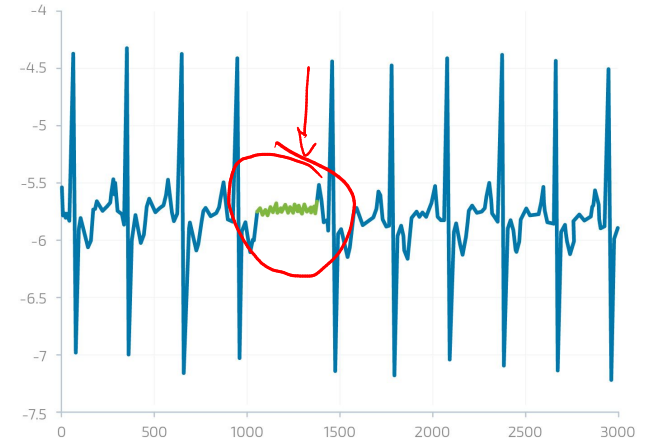
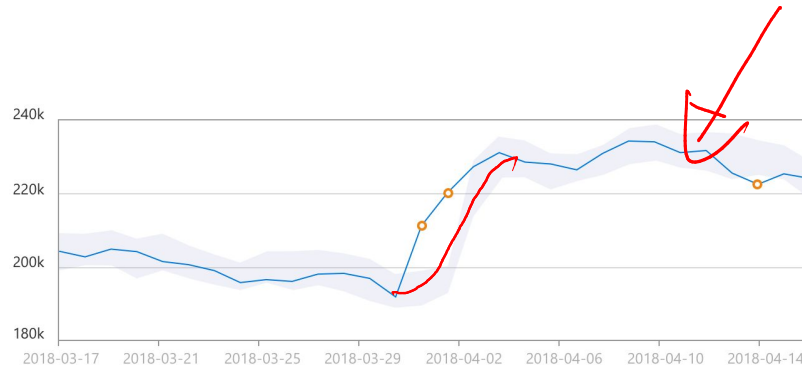
- Trepidação de máquina
- Pressão no duto
- Temperatura do equipamento
- ...

Esses fenômenos já tem um comportamento esperado a priori e uma anomalia é simplesmente detectada pela definição de limiares.

Anomalias



<https://medium.com/@DomaniNews/iot-data-anomaly-detection-using-machine-learning-75ab2296c4cc>



<https://www.infopulse.com/blog/a-new-bi-vector-of-real-time-and-ai-enabled-anomaly-detection>

<https://docs.microsoft.com/pt-br/azure/cognitive-services/anomaly-detector/overview>

Anomalias

Há uma infinidade de algoritmos para lidar com anomalias

- Um deles é calcular um erro esperado de um forecaster;
 - Fazer a previsão de forma janela na sua série
 - Verificar pontos ou subsequências fora da margem de erro

Nesse momento, eu vou passar reto. Mas, sobrando tempo, tentem fazer.

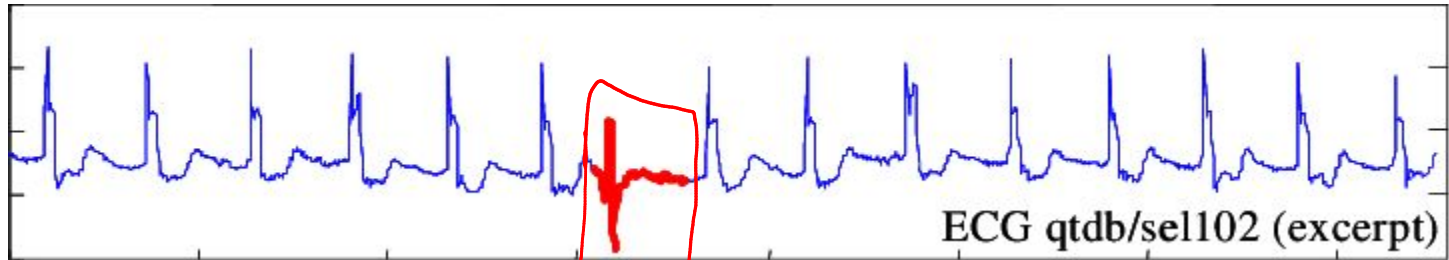
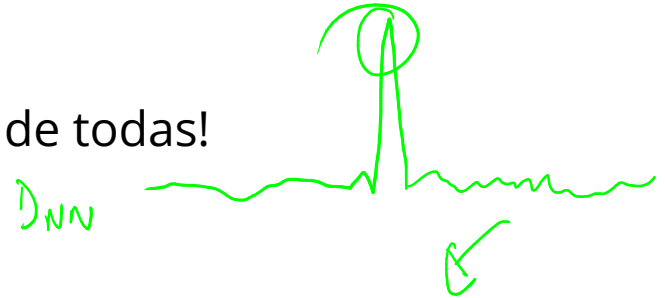
Anomalias

- Detecção online vs offline
- Early detection
- Individual/pontual vs coletiva
- Local vs. global

Anomalias

Um ponto de vista: a subsequência mais diferente de todas!

- Discord



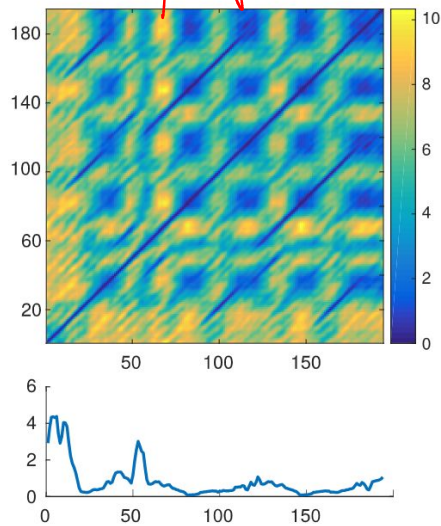
Discord

KEOGH, Eamonn et al. Finding unusual medical time-series subsequences: Algorithms and applications. IEEE Transactions on Information Technology in Biomedicine, v. 10, n. 3, p. 429-439, 2006.

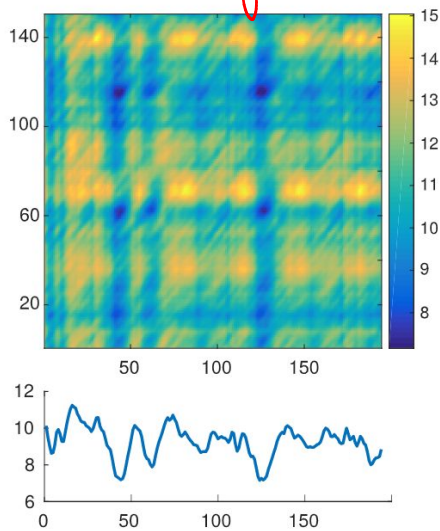
Matrix Profile (MP)

MP - Intuição

Self-joins

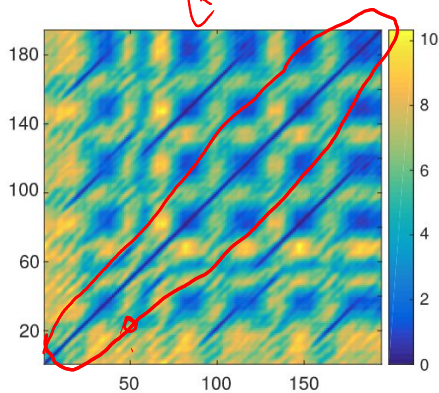


AB-joins



MP - Intuição

Self



subsequence
Self-sim. matrix

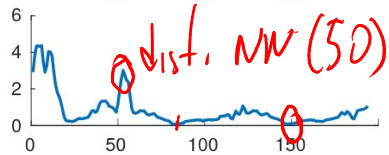
$SSM(50, 20) =$

$DE(x_{50 \dots 50+l}, x_{20 \dots 20+l})$

MP \rightarrow

$\in NN\text{-join}$

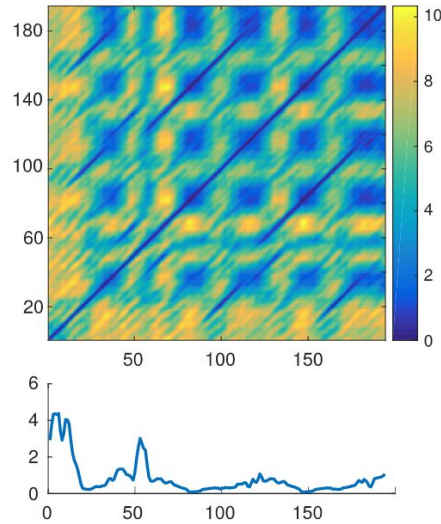
MP Index



$l = \text{comp. subseq.}$

\hookrightarrow preciso de uma zona de exclusão

MP - Intuição

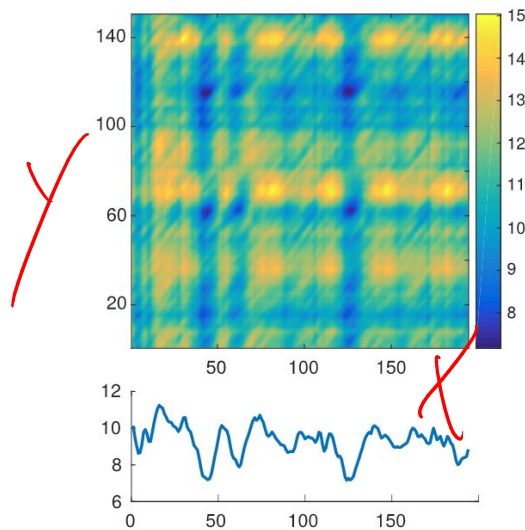


$$(A-B)^2 = \underbrace{A^2 + B^2 - 2A \cdot B}$$

z-normalização

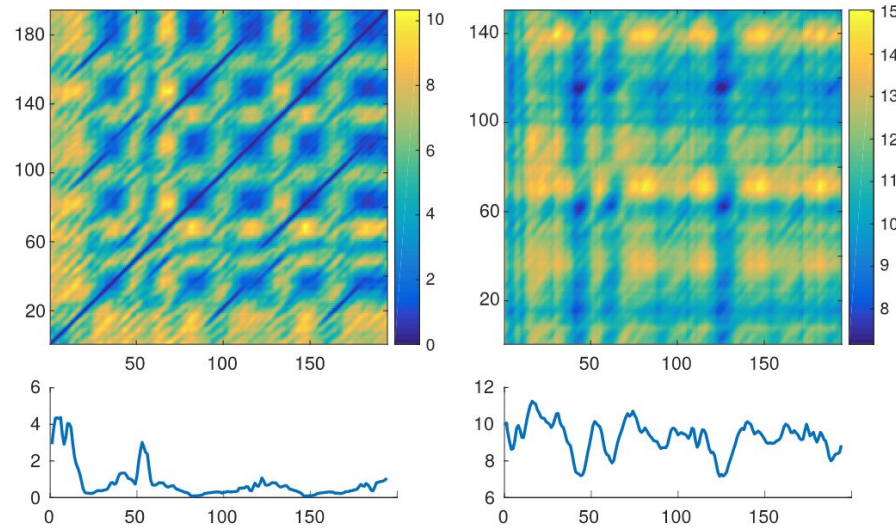
MP - Intuição

AB-join



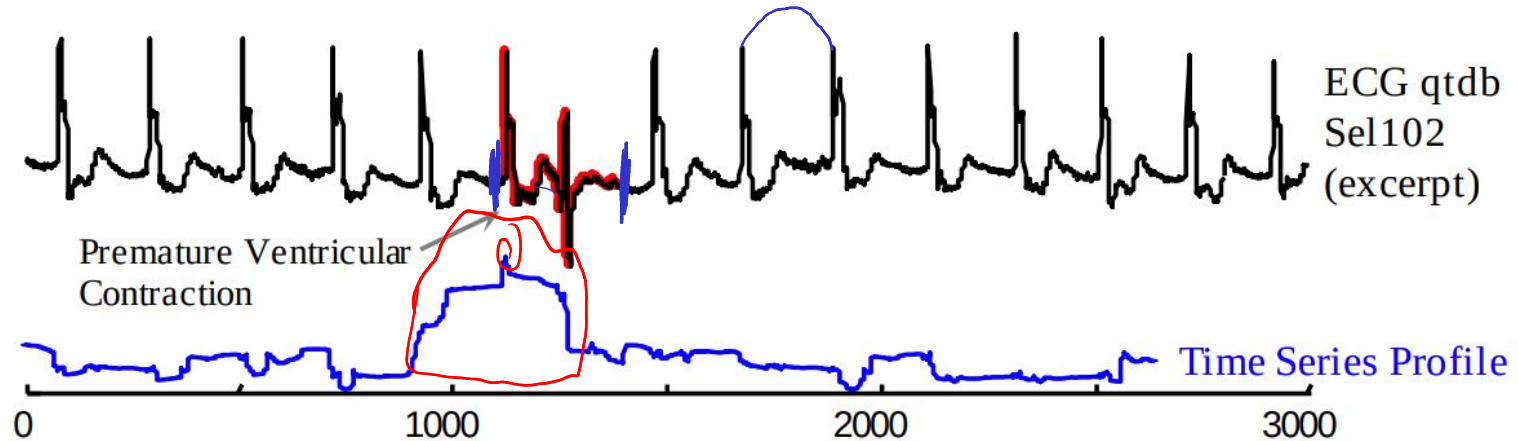
→ SIMPLE
P/
cover
song
ident.

MP - Intuição



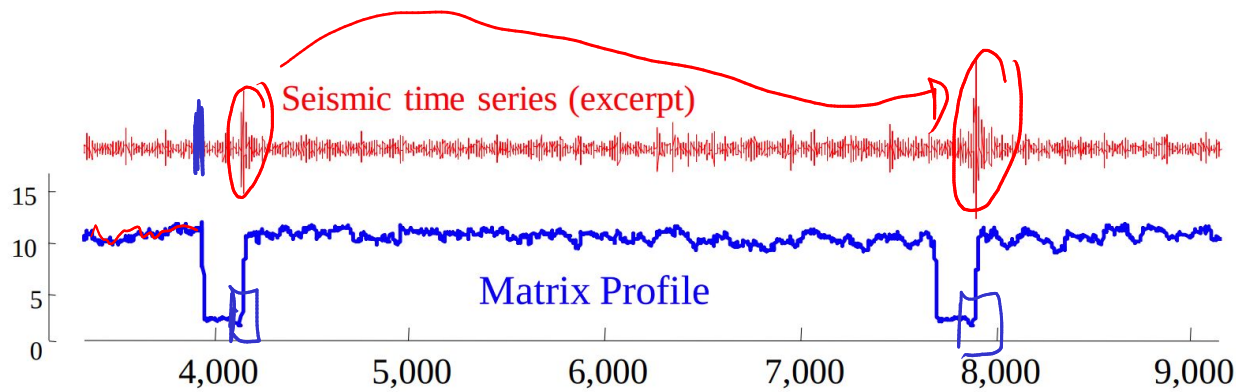
<- self-join e AB-join ->

MP - exemplo de discord

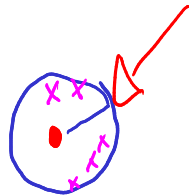


Yeh, C. C. M., Zhu, Y., Ulanova, L., Begum, N., Ding, Y., Dau, H. A., ... & Keogh, E. (2016, December). Matrix profile I: all pairs similarity joins for time series: a unifying view that includes motifs, discords and shapelets. In 2016 IEEE 16th international conference on data mining (ICDM) (pp. 1317-1322). IEEE.

MP - exemplo de motif → pair-based motif



Obs:
Support-based
motif



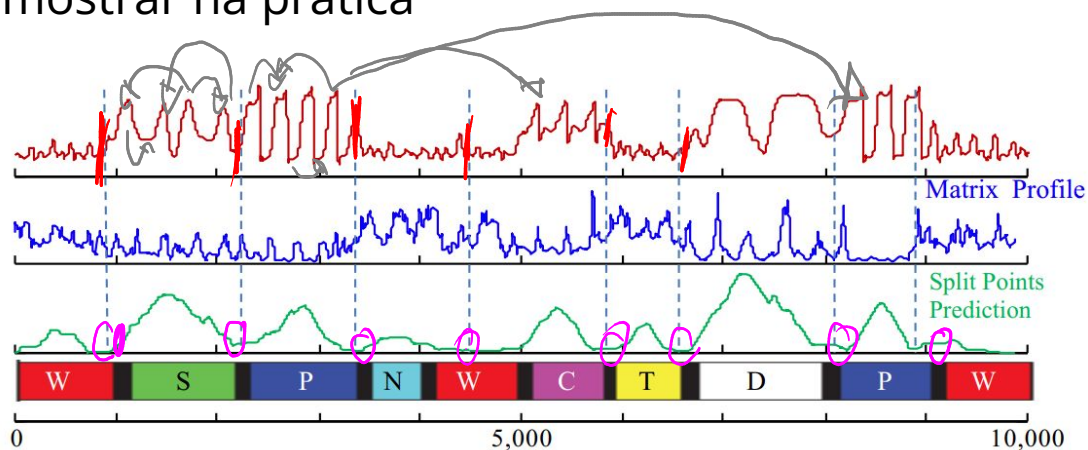
Yeh, C. C. M., Zhu, Y., Ulanova, L., Begum, N., Ding, Y., Dau, H. A., ... & Keogh, E. (2016, December). Matrix profile I: all pairs similarity joins for time series: a unifying view that includes motifs, discords and shapelets. In 2016 IEEE 16th international conference on data mining (ICDM) (pp. 1317-1322). IEEE.

MP - segmentação semântica

PEFA
↳ INPEFA

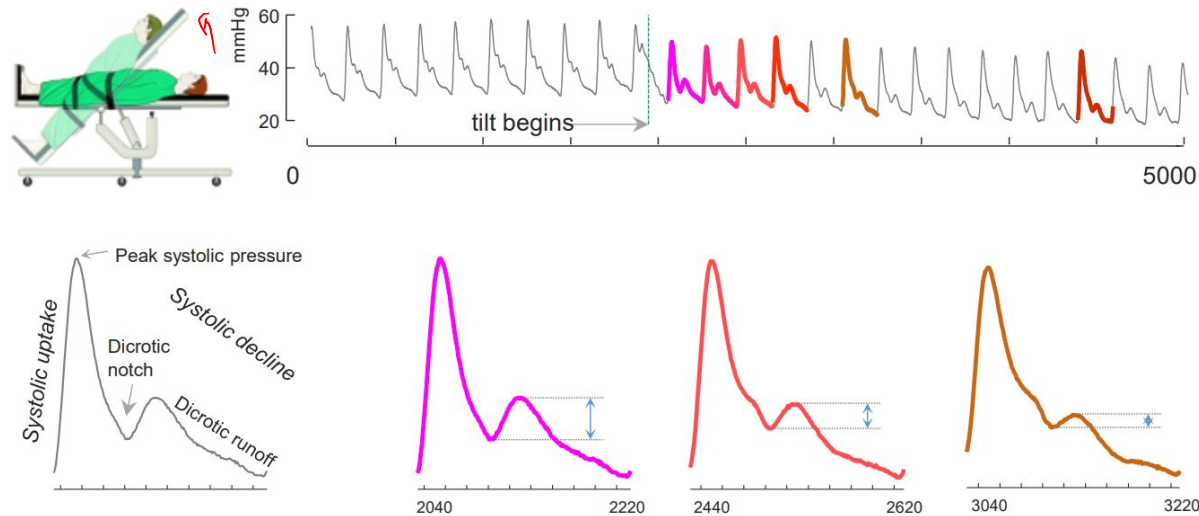
Aqui ainda era embrião

- O Luiz vai mostrar na prática



MP - time series chains

Padrões que evoluem no tempo



MP - pausa para a prática